Dynamical networks in tRNA:protein complexes

Anurag Sethi[a,1], John Eargle[b,1], Alexis A. Black[a], and Zaida Luthey-Schulten[a,b,2]

[a]Department of Chemistry, [b]Center for Biophysics and Computational Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Community network analysis derived from molecular dynamics simulations is used to identify and compare the signaling pathways in a bacterial glutamyl-tRNA synthetase (GluRS):tRNA^{Glu} and an archaeal leucyl-tRNA synthetase (LeuRS):tRNA^{Leu} complex. Although the 2 class I synthetases have remarkably different interactions with their cognate tRNAs, the allosteric networks for charging tRNA with the correct amino acid display considerable similarities. A dynamic contact map defines the edges connecting nodes (amino acids and nucleotides) in the physical network whose overall topology is presented as a network of communities, local substructures that are highly intraconnected, but loosely interconnected. Whereas nodes within a single community can communicate through many alternate pathways, the communication between monomers in different communities has to take place through a smaller number of critical edges or interactions. Consistent with this analysis, there are a large number of suboptimal paths that can be used for communication between the identity elements on the tRNAs and the catalytic site in the aaRS:tRNA complexes. Residues and nucleotides in the majority of pathways for intercommunity signal transmission are evolutionarily conserved and are predicted to be important for allosteric signaling. The same monomers are also found in a majority of the suboptimal paths. Modifying these residues or nucleotides has a large effect on the communication pathways in the protein:RNA complex consistent with kinetic data.

aminoacyl-tRNA synthetase | communication networks | community | suboptimal paths

In the modern world of translation, aminoacyl-tRNA synthetases (aaRSs) help maintain the genetic code by charging tRNA with its cognate amino acid. The formation of aminoacyl-tRNAs (aa-tRNAs) proceeds via a 2-step process. In the first step, the amino acid or its precursor reacts with ATP to form the activated aminoacyl-adenylate (aa-AMP) within the catalytic site, and in the second, or charging step, the amino acid is transferred to the 3′ end of the cognate tRNA. The aaRSs distinguish a particular set of tRNA species from a pool of many tRNA molecules in the cell through interactions with a group of nucleotides called the identity elements. For most aaRSs, the tRNA identity elements include the anticodon bases 34–36 and the discriminator base 73 in addition to other locations that are specificity dependent. In a few cases like leucyl-RS (LeuRS) in archaea, the synthetase has acquired additional domains that interact with identity elements on the variable arm of the tRNA instead of interacting with the anticodon (1).

Upon binding, the tRNA induces conformational changes throughout the protein:tRNA interface and within the catalytic site (2). Based on biochemical studies, the charging reaction is stimulated by interactions between the synthetase and the tRNA identity elements, which are mostly located far away from the site of amino acid attachment. Such long distance coupling is at the very heart of allosteric regulation (3). Experimental and computational studies of many regulatory complexes support the current view that they possess the intrinsic ability to undergo conformational transitions, conferred by the 3-dimensional network of interresidue interactions (4–8). The pathways of signal transduction favored by the network of interresidue contacts and the role conservation plays in these pathways remain to be established. This study demonstrates that nucleotides in the tRNA as well as residues within the aaRS are essential for information transduction in the protein:RNA complex. Although contact maps based on the static structure of the complex give an initial approximation to the physical communication network, the inclusion of dynamical correlations provides a more accurate picture of the network topology and approximates the strength of the allosteric signal that can be related to experimental observations.

For a given fold topology, contact maps generate unweighted networks representing the residue connectivity (9). The contribution of each residue or node to the characteristic path length (CPL), defined as an average of the shortest path length between all pairs of nodes in the network, provides an estimate of the effect of node connectivity on communication pathways in a protein. Conserved residues that greatly affect the CPL upon removal have been hypothesized to be important for allosteric signal transmission (10). Snapshots from a short simulation of a modeled MetRS:tRNA complex indicated that the shortest path between protein residues interacting with the anticodon and the adenylate binding site was sensitive to conformational changes in the protein (11), but the tRNA and contacts with other identity elements on the tRNA were neglected in their study of the signal transmission. Although the shortest path analysis identifies several nodes, the contribution of these nodes to communication in protein networks has not been examined, with few exceptions (12).

If there are multiple communication paths nearly equal in length, then not all residues along these paths need be considered as important for allostery. Instead, only residues or interactions that occur in the highest number of suboptimal pathways need to be conserved to guarantee an effective pathway for allosteric communication in the complex. In this work, we analyze entire protein:tRNA networks "weighted" by correlation data from long (20 ns) molecular dynamics (MD) simulations of the aaRS:tRNA complex in 2 functional states: before and after tRNA aminoacylation. The correlation, C_{ij}, in motion between nodes i and j defines information transfer between the nodes because motion of monomer (residue or nucleotide) i can be used to predict the direction of motion of monomer j. For all states, we determine the shortest path for communication along with the ensemble of suboptimal paths from all identity elements on the tRNA to the active site of the synthetase. The time-averaged connectivity of the nodes is used to identify the substructure or communities in the network. The optimal community distribution is calculated by using the Girvan–Newman algorithm (13), which has no free parameters, in contrast to other approaches (12, 14). The community description allows us to compare the topology and modularity of networks for the protein:tRNA complexes for 2 diverse class I aaRSs. The con-

Table 1. The shortest distance ($D_{i,A76}^0$) and number of suboptimal paths (N_{sop}) from each identity element to A76 in the networks representing the pre- and posttransfer states

	Pretransfer state		Posttransfer states	
Source	$D_{i,A76}^0$	N_{sop}	$D_{i,A76}^0$	N_{sop}
G1	4.31	3	13.31 (—)	2 (—)
C72	1.43	3	9.68 (—)	2 (—)
G2	3.29	31	7.45 (5.55)	3 (11)
U71	1.49	3	6.15 (3.45)	5 (11)
C4	3.04	85	8.00 (2.51)	29 (79)
G69	4.19	75	5.95 (2.90)	1 (75)
U11	4.88 (4.91)	209	8.89 (3.13)	30 (85)
A24	4.0 (4.11)	177	7.80 (4.00)	40 (238)
U13	3.24 (5.19)	104	6.23 (2.77)	29 (81)
G22	4.14 (5.03)	105	6.74 (4.11)	29 (100)
A46	5.13 (6.02)	106	8.14 (3.71)	36 (88)
C34	6.45	315	11.05 (5.21)	83 (212)
U35	5.48	196	9.95 (4.56)	84 (177)
C36	5.10	204	10.85 (4.53)	60 (171)
A37	4.71	230	9.73 (4.14)	60 (215)

The values in parentheses in the pretransfer state (GluRS:tRNA^{Glu}:Glu-AMP) denote the distance of the identity element from A76 in the modified system. The values in the posttransfer network denote the shortest distance and number of suboptimal paths for the posttransfer network in 2 states: GluRS:Glu-tRNA^{Glu}:AMP and GluRS:Glu-tRNA^{Glu}:HAMP (in parentheses).

Fig. 1. Correlation analysis (C_{ij}) of the motion during a 20-ns MD simulation of the GluRS complex. Monomers with highly (anti)correlated motion are orange or red (blue). Distant (>15Å) regions displaying high degree of (anti-)correlation are marked in white rectangles (below)above the diagonal.

served monomers involved in communication between communities are the critical nodes for communication within the network and are shown to occur in a majority of the suboptimal paths between the identity elements and the site of amino acid transfer at the 3′ end of the tRNA.

The aaRSs are multidomain proteins that are divided into 2 classes based on the homology of their catalytic domain that catalyzes both steps of the aminoacylation reaction. Most of the class I aaRSs are monomeric enzymes, whereas the class II aaRS enzymes form dimers or tetramers in solution. Because of the smaller size and simpler nature of the functional aaRS:tRNA complex in the class I aaRS family, we investigate allostery in these complexes. The different modes of tRNA recognition by GluRS and LeuRS offer insight into the evolution of the allosteric network upon changes in specificity.

Results and Discussion

All members of the class I aaRS family have homologous catalytic domains formed from the Rossmann fold and a specificity-dependent connective polypeptide 1 (CP1) insertion as shown in Fig. S1. The evolution of different (tRNA and amino acid) specificities in the aaRSs proceeded via the acquisition of additional domains, which in some cases resulted in different patterns of interaction between the tRNA and aaRS. In GluRS and LeuRS, the catalytic domain is followed by α-helical and C-terminus domains (CTD) that interact with the tRNA. In GluRS, the CTD interacts with the anticodon loop and is called the anticodon-binding (ACB) domain, whereas in archaeal LeuRS, residues in the CTD form contacts with the long variable arm and the elbow region of the tRNA. GluRS and LeuRS interact differently with their cognate tRNA molecules, resulting in vastly different identity elements. Although GluRS makes contact with the identity elements listed in Table 1 on the acceptor stem, GG or D arm, and the anticodon loop of tRNA^{Glu}, the identity elements of the archaeal LeuRS are in the long variable arm and discriminator base of the tRNA (see Table S1). The bacterial GluRS involved in the direct pathway for glutamate aminoacylation (discriminate GluRS or D-GluRS) is

investigated in this work and later compared with the network for the archaeal version of LeuRS.

Correlation Analysis. The transmission of an allosteric signal within the protein:RNA complex should couple motion between active site residues and regions in the protein interacting with identity elements on the tRNA. The Rossmann fold forms the active site for the aminoacylation reaction and interacts with identity elements on the acceptor stem and the GG arm. In Fig. 1, the degree of coupled motion in the GluRS:tRNA^{Glu}:Glu-AMP (pretransfer) complex was measured by normalizing the cross correlation matrix of atomic fluctuations over the length of the simulation.

Besides local correlations, there is coupling between distant parts of the complex shown as boxed regions in Fig. 1. Motion of the α-helical ACB domain is anticorrelated to that of the Rossmann fold. Similarly, the dynamics of the CP1 insertion are coupled to the dynamics of the Rossmann fold, the 4-helix junction (4HJ), the ACB domain, and the anticodon loop. Most significantly, the C-terminus half of the Rossmann fold is dynamically correlated to the motion of the anticodon, despite these regions being 55 Å apart. Although the longer simulations provided a more pronounced correlation map, the trends are similar to those observed in shorter simulations of the MetRS:tRNA^{Met} complex (11). The simulation of the archaeal LeuRS:tRNA^{Leu} complex displays similar coupled motion between distant regions in the complex as shown in Fig. S2. The long-range coordinated motion for the pretransfer complex is also observed in the 3 most dominant principal components of the MD simulation (Figs. S3 and S4).

Although correlation analysis provides evidence for the presence of allostery, the communication pathways between various regions of the complex cannot be elucidated by using solely these methods. The pathways and the residues/nucleotides in the protein:RNA complex critical for communication have been determined from network methods as described below.

Each residue and nucleotide in the protein:RNA complex represents a node in the network. Any 2 nonneighboring monomers are connected by an edge if they are in contact during a majority of the simulation. In the dynamic network, the edges are weighted by the correlation values from the simulation so that the distance between 2 nodes connected by an edge reduces as the correlation (or energy of interaction) between the monomers increases.
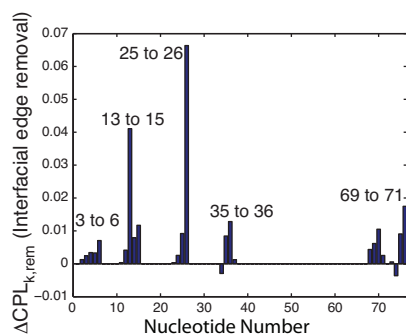
Fig. 2. Difference in characteristic path length: the change in CPL upon edge removal for each nucleotide at the interface of tRNA^{Glu} in the GluRS network. The nucleotides with significant increase in CPL are labeled.

Characteristic Path Length Analysis. Allosteric signal transmission in the aaRS:tRNA complex involves communication of dynamical information within both macromolecules. The interface of the protein:RNA complex initiates the signal for amino acid transfer to the tRNA. To identify nucleotides that have the largest effect on communication across the interface, the change in CPL is calculated upon removing all contacts from a given interface nucleotide to any residue on the protein while keeping all other contacts in the network intact. Nucleotides that significantly increase the edge CPL in Fig. 2 are either at or close to G26 or the identity elements located on the acceptor stem, GG arm, and the anticodon arm. Besides physically connecting the GG and anticodon arms, nucleotides 25 and 26 are important for coordinating communication between the identity elements in these 2 arms.

Change in CPL upon the complete removal of a node is a measure of its effect on communication within the entire network. Although several residues and nucleotides increase the CPL upon their removal from the pretransfer network (see Figs. S5 and S6 and Table S2), the results are difficult to interpret because this global metric is sensitive to the geometry of the network, and it underestimates the contribution of nodes at the periphery of the network.

Community Analysis of tRNA:GluRS. Variations in the connectivity of the network give rise to modules or local communities in the network. Analysis of the communication between these local substructures provides additional insight not readily available through the global CPL measurement. Nodes belonging to the same community are more strongly and densely interconnected to one another and have weaker connections to other nodes in the network. By definition, nodes in the same community can communicate with one another relatively easily through multiple routes. However, there are comparatively few edges involved in communication between communities, and the monomers involved in this communication form a bottleneck for information transfer in the network.

The Girvan–Newman algorithm splits the network of GluRS:tRNA^{Glu}:Glu-AMP into 13 communities as shown in Table S3 and Fig. 3. These communities do not necessarily correspond to the domain definitions of the protein and the RNA. Monomers in the same community in Fig. 3 are local in structure but can be distant in sequence. Rearranging the cross-correlation map by communities in Fig. S7 clearly shows that monomers within the same community are highly correlated. Most of the interface nucleotides (including the identity elements) are found in the same community as their protein interaction partners. Of the 13 communities, there are 3 communities composed exclusively of tRNA nucleotides, 8 communities including a combination of tRNA and protein monomers,

and 2 communities containing only protein residues. The main signal for allostery is assumed to travel from the communities with the various identity elements, i.e., communities C-3, C-4 (anticodon loop), C-8 (GG arm), C-1, C-6, and C-7 (acceptor stem), to the active site of the aaRS, which is at the interface of C-6 (A76) and C-2 (adenylate Glu-AMP). The nodes representing nucleotides U20A and U59 are not connected to any other nodes in the network, form isolated communities of their own, and are not shown in Fig. 3C. If nearest-neighbor interactions were allowed in this network, both nucleotides would be merged into the community containing their neighboring nucleotides (C-9).

The flow of information in the physical network of the protein:RNA complex is traced by using the coarse-grained picture formed by the network of communities shown in Fig. 3C. The betweenness of an edge, defined as the number of shortest paths that pass through the edge in the network, is used to measure the importance of the edge for communication within the network. The width of an edge connecting 2 communities in the community network is proportional to the sum of betweenness of edges connecting them in the protein:RNA network. The adenylate/Rossman fold community (C-2) is central to the information flow, connecting the 4HJ and ACB domain on one side to the catalytic domain (C-1, C-7, and C-11) and the tRNA region (C-8) on the other. The communities spanning the GG arm and the anticodon arm, C-8 and C-10, are weakly connected and suggest that information flows through the tRNA in addition to the synthetase. This is supported by the study of the tRNA^{Val} split into 2 minihelices where the anticodon minihelix stimulates aminoacylation of the acceptor minihelix in ValRS (15). The monomers that occur in a majority of the (shortest) pathways for intercommunity information transfer are listed in Table S4 and are predicted to be critical for allostery. A large number of conserved residues close to the Glu-AMP binding site are important for intercommunity communication (Pro 8, Tyr 20, Ile 56, Pro 228, His 232, and Pro 234). In addition, conserved protein residues (Gly 274 and Phe 305) close to or interacting with identity elements on the GG arm (U11 and A14) are also predicted to be necessary for allosteric signal transfer. Leu 359 was identified as important for communication between the 4HJ and the ACB domain, and although it is not highly conserved, its neighbor Arg 358 has been shown to play a crucial role in anticodon recognition by GluRS and its ability to distinguish tRNA^{Gln} and tRNA^{Glu}. Comparison of the monomers identified by the CPL and community analyses of dynamic and static protein:RNA networks reveals a major difference between the methods. In general, the community analysis identifies far fewer critical monomers than the CPL analysis of the dynamic and static protein:RNA networks (25 versus 54 and 47, respectively). The CPL analysis selects nodes that occur in a large number of paths for both intra- and intercommunity communication. Nodes within the same community are highly interconnected and can communicate through a large number of paths with a small difference in distance. Hence, the nodes that are important for intracommunity communication have a smaller effect on communication throughout the network, as evidenced by their lower conservation. Based on Tables S2 and S4, the set of monomers identified from the community analysis has 85.8% conservation on average whereas those identified by the CPL analyses of the dynamic and static networks have a mean conservation of 67.6% and 63.4%, respectively.

Modifications Alter the Network. Mutation of the identity element U13 in the GG arm reduces the catalytic efficiency of D-GluRS to aminoacylate its cognate tRNA by a factor of 50 in *Escherichia coli* (16). Computationally, without carrying out another MD simulation, this modification was captured by weakening the edges between U13 and any residue on the synthetase in the network analysis. Weakening the interface edges of U13 or its

Fig. 3. Community analysis of the network formed based on the GluRS adenylate simulation. (*A* and *B*) The monomers are colored (*A*) according to community membership calculated in *B*: cyan, 1; purple, 2; orange, 3; green, 4; lime, 5; blue, 6; tan, 7; black, 8; yellow, 9; red, 10; and ochre, 11. Hard spheres indicate residues that occur in a majority of shortest paths connecting nodes in different communities. The width of the lines is proportional to the betweenness of the edge (number of shortest paths passing through that edge). (*C*) Community network representation: The width of the lines is proportional to the number of shortest paths passing through those junctions, and the presence of an identity element in a community is indicated by an asterisk. (*D*) Community network for modified system in which all contacts between U13 and GluRS are weakened. The isolated communities made of the single nucleotides (U20A and U59) are not shown in the community networks.

neighbor A14 to the protein leads to significant repartitioning among the community network (Fig. S8). These edges are removed early in the Girvan–Newman algorithm and hence have the largest overall effect on the community node assignment. The community network after reducing the correlation between U13 and GluRS by one-half is shown in Fig. 3*D*. As measured by the community repartition difference, 65% of the node pairs remain grouped in the same community, whereas 35% of the node pairs are split into separate communities. The boundaries of all of the communities vary slightly from the community distribution in the wild-type network. The most significant changes occur close to C-8 where a new community C-12 contains most of the strong contacts that C-10 previously formed with C-3 (U11 and C25) and C-8 (C25 and Trp 312). In the modified network, C-8 acquires residues and nucleotide C12 from C-3 and C-7, and, as a result, the edge between C-7 and C-3 is replaced with the edge between C-8 and C-3. A similar study for the computational alanine scan of protein residues at the interface is shown in Fig. S9.

The increase in the shortest distance (Table 1) and CPL (Fig. 2) ($CPL_{wt} < CPL_{mut}$) upon complete removal of the interface contacts of U13 corresponds to a lower experimentally determined efficiency [$(k_{cat}/K_M)_{wt} \gg (k_{cat}/K_M)_{mut}$] (16) and weaker allosteric signal in the mutated complex. The overall sum of shortest distances between the identity elements and A76 in the modified network increases by 3.89 compared with the wild-type network. This implies that the probability of informa-

tion transfer along these paths, the product of correlations, differs by a factor of $e^{3.89} = 47.9$. In GluRS, the tRNA is required for both steps of aminoacylation, but only the amino acid transfer step is modeled in this study. It may be fortuitous that the decrease in probability of information transfer in the modified network agrees so well with the drop in catalytic efficiency.

Comparison of Pretransfer and Posttransfer Networks. The network representing the system after tRNA aminoacylation is obtained from a simulation of the posttransfer complex in 2 different states—GluRS:Glu-tRNA^{Glu}:AMP and GluRS:Glu-tRNA^{Glu}:HAMP. In the proposed mechanism for the homologous GlnRS (17), the phosphate of Gln-AMP plays the role of a general base and a proton is transferred from the 2′-hydroxyl of A76 in the tRNA to the phosphate with the concomitant formation of singly protonated AMP (HAMP). This proton could then be transferred to a histidine in the HIGH motif or to the solvent molecules in the catalytic pocket to form GluRS:tRNA^{Glu}:Glu-AMP.

To compare the pre- and posttransfer states, the shortest distance and the suboptimal paths between the experimentally determined identity elements (16, 18) and A76 at the 3′ end of the tRNA were measured in the networks and are reported in Table 1. As the charging amino acid is attached to A76 in the aminoacylation process, this nucleotide serves as the target for the transmission of the allosteric signal. The shorter the distance, the larger the correlation of the monomers along the path in the network and the greater the allosteric signal in the protein:RNA
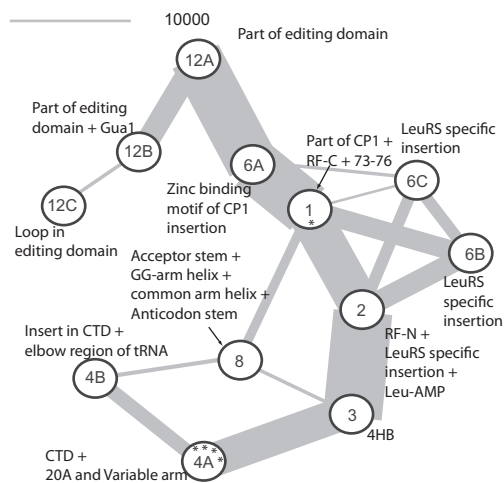
Fig. 4. Scaled community network of the LeuRS:tRNA adenylate complex.

complex (see *Methods*). The shortest distances are always displayed by the pretransfer and the GluRS:Glu-tRNA^{Glu}:HAMP posttransfer state. This indicates that the correlations are larger in the complex that is modeled closest to the transition state and decrease as the substrates begin to undock.

A large number of suboptimal paths in the pretransfer network means that there are many alternative paths in the network with nearly equivalent distance. This implies that not all residues along the path are necessary for signaling, but only a few nodes that occur in a majority of these suboptimal paths are critical for allostery. These nodes are listed in Table S5, and approximately half of them are identified in the community analysis. In addition, several other nodes that appear in the suboptimal paths are sequence or structural neighbors to nodes appearing in the community analysis. From this comparison, we conclude that the intercommunity junctions are crucial regions for the communication of allosteric signal between A76 and the identity elements.

Network Analysis of tRNA:LeuRS. The mode of tRNA^{Leu} recognition by LeuRS is dramatically different from that of other class I aaRSs. The anticodon loop no longer makes direct contact with the protein, and instead, tRNA^{Leu} has evolved a long variable arm that interacts with the CTD of LeuRS. Identity elements on the variable arm replace those in the anticodon loop and stem used in other tRNAs, which allows LeuRS to recognize up to 6 different tRNA isoacceptor species inside the cell. Repartition of communities upon modification of the interface edges (Fig. S10) clearly demonstrates the shift in recognition to the variable arm of the tRNA and the loop in the GG arm. In addition, LeuRSs have evolved an editing domain in the middle of the CP1 insertion that deacylates misaminoacylated tRNA. These differences in recognition have an impact on the physical network topology of the LeuRS:tRNA^{Leu} complex even though the core of the community network remains the same.

There are 11 major communities in the pretransfer network as listed in Table S6. Similar to the GluRS network, the Rossmann fold forming the core of the catalytic domain in LeuRS splits into 2 communities made from the N-terminus half (C-2) and the C-terminus half of the Rossmann fold (C-1) as shown in Fig. 4. The 4HB (C-3) interacts with the anticodon stem and is comparable with the community of the 4HJ (C-3) in the GluRS network. The CTD interacts strongly with identity elements in the variable arm and with the GG loop, forming 2 communities (C-4A and C-4B) that are analogous to the ACB domain of the GluRS network (C-4). The larger CP1 insertion (previously C-5 and C-6) forms independent communities in the network (C-6A,

C-6B, and C-6C). Of these, C-6A in the LeuRS system is topologically equivalent to C-6 in the GluRS network. The acquired editing domain forms 3 additional communities (C-12A, C-12B, and C-12C). The tRNA communities C-8, C-9, and C-10 in the GluRS network unite into a single community in the LeuRS network (C-8). Structural overlap of the 2 synthetase complexes clearly shows that the acceptor stem on tRNA^{Leu} now makes contact with the C-terminus of the Rossmann fold, which establishes the network connections from C-8 to C-1.

In the pretransfer complex for LeuRS, the main signal for allostery is from the tRNA variable arm (C-4A) and the discriminator base (C-1) to A76 and the adenylate (interface of C-1 and C-2) in the active site. Similar to the GluRS system, residues identified in the community network analysis of the LeuRS system are also highly conserved as shown in Table S7, but are not conserved across different specificities, except for the HIGH motif. Some of these residues could also play a role in the specificity of the aaRS enzyme.

Conclusion

The communication pathways that lead to coordinated motion between functionally important and distant regions of the protein:RNA complexes are highly degenerate. In these degenerate pathways, only a few nodes that occur at intercommunity junctions control the communication within the complex. These nodes also appear in the majority of the suboptimal paths between the identity elements on the tRNA and the active site in the synthetase and are predicted to be important for allostery. The community picture provides a coarse-grained view of the network that can be used to compare topologically similar aaRS:tRNA networks even when some of the domains are structurally unrelated. The core of both class I aaRS:tRNA networks compared in this study is formed by the 2 communities made from the Rossmann fold, whereas equivalent roles are provided for RNA-binding domains that have evolved later in these enzymes. Our analysis of the dynamical networks provides several metrics for comparing the signaling in different states and/or modifications of the systems and is applicable to other protein:RNA and protein:protein complexes.

Methods

Weighted RNA:Protein Network. A network is defined as a set of nodes with connecting edges. Amino acid residues, nucleotides, and the AMP substrate are each represented by a single node. Edges connect pairs of nodes if the corresponding monomers are in contact, and 2 nonconsecutive monomers are said to be in contact if any heavy atoms (nonhydrogen) from the 2 monomers are within 4.5 Å of each other for at least 75% of the frames analyzed. Changes in the parameters defining the network contacts lead to minor changes in the community distribution of the network (Fig. S11).

Nearest neighbors in sequence are not considered to be in contact as they lead to a number of trivial suboptimal paths in the weighted network. The

dynamical networks are constructed by using data from the final 16 ns of 20-ns trajectories of the protein:RNA complexes sampled every 50 ps. The dynamical protein:tRNA network is a weighted network in which the weight (w_{ij}) of an edge between nodes i and j is the probability of information transfer across that edge as measured by the correlation values between the 2 monomers in the simulation: $w_{ij} = -\log(|C_{ij}|)$. This definition gives a probabilistic interpretation of the lengths of shortest paths and treats strong correlations and anticorrelations similarly (Fig. 1 and Fig. S2).

Shortest Paths, Betweenness, and Suboptimal Paths. The length of a path D_{ij} between distant nodes i and j is the sum of the edge weights between the consecutive nodes (k,l) along the path: $D_{ij} = \sum_{k,l} w_{kl}$. The shortest distance D_{ij}^0 between all pairs of nodes in the network is found by using the Floyd–Warshall algorithm. The betweenness of an edge is the number of shortest paths that cross that edge. The average of all shortest paths, known as the CPL, is a measure of the network size.

Although the shortest path is the most dominant mode of communication between the nodes, the number of paths within a certain limit δ of the shortest distance is a measure of the path degeneracy in the network. All suboptimal paths for communication between the active site and the identity elements are determined in addition to the shortest path. The tolerance value used for any alternate path to be included in the suboptimal path was $-\log(0.5) = 0.69$, which is close to the average protein edge weight. The trends shown in Table 1 remain the same for cutoffs of δ = 0.25 and 0.1. On average, with δ = 0.5, about 15% of the paths in Table 1 traverse the same node twice during a single suboptimal path.

Community Analysis. The physical network of nodes and edges contains substructures or communities of nodes that are more densely interconnected to each other than to other nodes in the network. The community structure is identified by using the Girvan–Newman algorithm (13), which uses a top-down approach to iteratively remove the edge with the highest betweenness and recalculate the betweenness of all remaining edges until none of the edges remain.

The optimum community structure is found by maximizing the modularity value Q, which is a measure of difference in probability of intra- and inter-community edges. Q can have a maximum value of 1; large values of Q indicate better community structure. As the algorithm divides the network into increasingly smaller communities, the modularity score is measured for each community division, and the maximum value corresponds to the optimal community distribution of the network. In networks based on the 3D structure of the protein:tRNA complex presented here, the optimal modularity score is found to be ≈0.7. In typical real world networks, the optimal modularity score is in the range of 0.4–0.7 (29). More recently, a number of algorithms have been developed that explore different strategies for dividing a network into community structures, but they are more complex (30, 31).

Information Paths and Community Identification of Residues Important for Allostery. The shortest paths between pairs of nodes belonging to 2 different communities are calculated and analyzed for communication across communities in the network. Of these intercommunity links, all edges connecting any 2 of these communities are identified. Edges with the greatest betweenness are pinpointed, and the nodes connected by these edges are established as critical for allosteric signal transduction.

The strength of allosteric signal (A) is defined in this work as indirectly proportional to the sum of the shortest distances from the identity elements to A76:

$$A = \frac{1}{\sum_i D_{i,A76}^0}.$$

This value can be used to compare the strength of the allosteric signal between the wild-type enzyme in different states and/or modifications of the network.

1. Ibba M, Francklyn C, Cusack S, eds (2005) *The Aminoacyl-tRNA Synthetases* (Landes Bioscience, Georgetown, TX).
2. Sekine S, et al. (2003) ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J* 22:676–688.
3. Changeux J, Edelstein S (2005) Allosteric mechanisms of signal transduction. *Science* 308:1424–1428.
4. Miyashita O, Onuchic J, Wolynes P (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci USA* 100:12570–12575.
5. Hyeon C, Lorimer G, Thirumalai D (2006) Dynamics of allosteric transitions in GroEL. *Proc Natl Acad Sci USA* 103:18939–18944.
6. Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* 17:633–640.
7. Amaro RE, Sethi A, Myers RS, Davisson VJ, Luthey-Schulten ZA (2007) A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry* 46:2156–2173.
8. Hammes-Schiffer S, Benkovic S (2006) Relating protein motion to catalysis. *Annu Rev Biochem* 75:519–541.
9. Aszói A, Taylor W (1993) Connection topology of proteins. *Comput Appl Biosci* 9:523–529.
10. del Sol A, Fujihashi H, Amoros D, Nussinov R (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol Syst Biol* 2:1–12.
11. Ghosh A, Vishveshwara S (2007) A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc Natl Acad Sci USA* 104:15711–6.
12. Chennubhotla C, Bahar I (2006) Markov propagation of allosteric effects in biomolecular systems: Application to GroEL-GroES. *Mol Syst Biol* 2:36.
13. Girvan M, Newman M (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99:7821–7826.
14. Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435:814–818.
15. Frugier M, Florentz C, Giegé, R (1992) Anticodon-independent aminoacylation of an RNA minihelix with valine. *Proc Natl Acad Sci USA* 89:3990–3994.
16. Sekine S, et al. (1996) Major identity determinants in the ''augmented D helix'' of tRNA^Glu from *Escherichia coli. J Mol Biol* 256:685–700.
17. Perona JJ, Rould MA, Steitz TA (1993) Structural basis for transfer RNA aminoacylation by *Escherichia coli* glutaminyl-tRNA synthetase. *Biochemistry* 32:8758–8771.
18. Sekine S, Nureki O, Tateno M, Yokoyama S (1999) The identity determinants required for the discrimination between tRNA^Glu and tRNA^Asp by glutamyl-tRNA synthetase from *Escherichia coli. Eur J Biochem* 261:354–360.
19. Fukunaga R, Yokoyama S (2005) Crystal structure of leucyl-tRNA synthetase from the archaeon *Pyrococcus horikoshii* reveals a novel editing domain orientation. *J Mol Biol* 346:57–71.
20. Cusack S, Yaremchuk A, Tukalo M (2000) The 2 A crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J* 19:2351–2361.
21. Eswar N, Eramian D, Webb B, Shen M, Sali A (2008) Protein structure modeling with MODELLER. *Methods Mol Biol* 426:145–159.
22. Foloppe N, MacKerrell AD, Jr (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem* 21:86–104.
23. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802.
24. Eargle J, Black AA, Sethi A, Trabuco LG, Luthey-Schulten Z (2008) Dynamics of recognition between tRNA and elongation factor Tu. *J Mol Biol* 377:1382–1405.
25. Glykos NM (2006) Software news and updates. CARMA: A molecular dynamics analysis program. *J Comput Chem* 27:1765–1768.
26. Sethi A, O'Donoghue P, Luthey-Schulten Z (2005) Evolutionary profiles from the QR factorization of multiple sequence alignments. *Proc Natl Acad Sci USA* 102:4045–4050.
27. Roberts E, Eargle J, Wright D, Luthey-Schulten Z (2006) MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7:382.
28. Humphrey W, Dalke A, Schulten K (1996) VMD—Visual molecular dynamics. *J Mol Graphics* 14:33–38.
29. Newman M, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69:026113.
30. Newman M (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103:8577–8582.
31. Rosvall M, Bergstrom C (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci USA* 105:1118–1123.

CHEMISTRY

BIOPHYSICS AND COMPUTATIONAL BIOLOGY